

The Degradome database: expanding roles of mammalian proteases in life and disease

José G. Pérez-Silva, Yaiza Español, Gloria Velasco and Víctor Quesada*

From the Departamento de Bioquímica y Biología Molecular, Facultad de Medicina, Instituto Universitario de Oncología, Universidad de Oviedo, 33006 Oviedo, Spain

Received September 14, 2015; Revised October 23, 2015; Accepted October 26, 2015

ABSTRACT

Since the definition of the degradome as the complete repertoire of proteases in a given organism, the combined effort of numerous laboratories has greatly expanded our knowledge of its roles in biology and pathology. Once the genomic sequences of several important model organisms were made available, we presented the Degradome database containing the curated sets of known protease genes in human, chimpanzee, mouse and rat. Here, we describe the updated Degradome database, featuring 81 new protease genes and 7 new protease families. Notably, in this short time span, the number of known hereditary diseases caused by mutations in protease genes has increased from 77 to 119. This increase reflects the growing interest on the roles of the degradome in multiple diseases, including cancer and ageing. Finally, we have leveraged the widespread adoption of new webtools to provide interactive graphic views that show information about proteases in the global context of the degradome. The Degradome database can be accessed through its web interface at <http://degradome.uniovi.es>.

INTRODUCTION

Proteases catalyze the hydrolysis of peptide bonds in a fundamentally irreversible reaction. This means that these enzymes must be tightly regulated in terms of activation and specificity to avoid massive homeostatic disorders. In turn, this need for specificity has led to the evolutionary expansion of protease genes to regulate the correct proteolysis of a large set of substrates. The parallel expansion of protease inhibitor genes added an additional level of complexity to this biochemical process. In recognition of this complex and interwoven system, the degradome of an organism was defined as the complete set of proteases in that organism (1). The degradome has been shown to affect most of the characterized biochemical pathways. Thus, different proteases are known to play key roles in such biological processes

as cell cycle progression, tissue remodelling, neuronal outgrowth, haemostasis, wound healing, immunity, angiogenesis and apoptosis (2–6). Conversely, failures in the regulation of the degradome underlie diverse pathological conditions, including cancer, arthritis, progeria and neurological diseases (7–10).

Since the definition of the degradome relies on a global appraisal of the proteolytic processes, it follows that degradomics, the set of techniques specifically aimed at characterizing the degradome, must manage and integrate high-throughput data. In this regard, the completion of multiple genome projects allowed researchers to extend the degradomes of several species *in silico* from known protease sequences (11,12). Our experience in degradomics led us to tackle this problem with methods that relied heavily on manual curation after automatic predictions (13). As an additional advantage of this approach, a part of these projects consisted in the mining of the literature looking for known relationships between protease alterations and hereditary diseases, termed degradomopathies.

With this information, we described the Degradome database, containing the results of the manual annotation of every protease gene in the genomes of human, chimpanzee, mouse, and rat, along with relationships between protease alterations and hereditary diseases (14). This database complemented existing databases devoted to proteases, by providing a different focus. For instance, CutDB documents actual and predicted proteolytic events, but does not provide a global view of the proteases themselves (15). Also, MEROPS is a comprehensive and excellent database which relies on large-scale experiments and automatic predictions (16). By contrast, the Degradome database relies on manual annotation and exhaustive curation of genes, in multiple cases supported by direct cloning and sequencing experiments. For this reason, multiple instances of non-functional gene expansions, which our approach has filtered out, are annotated as putative proteases in the MEROPS database. In fact, the number of putative human proteases according to MEROPS is 990, whereas the Degradome database describes 588. In exchange, the MEROPS database features a very large number of species. Finally, our emphasis in diseases adds important information on the pathological rel-

*To whom correspondence should be addressed. Tel: +34-985 105025; Fax: +34 985 103564; Email: quesadavictor@uniovi.es

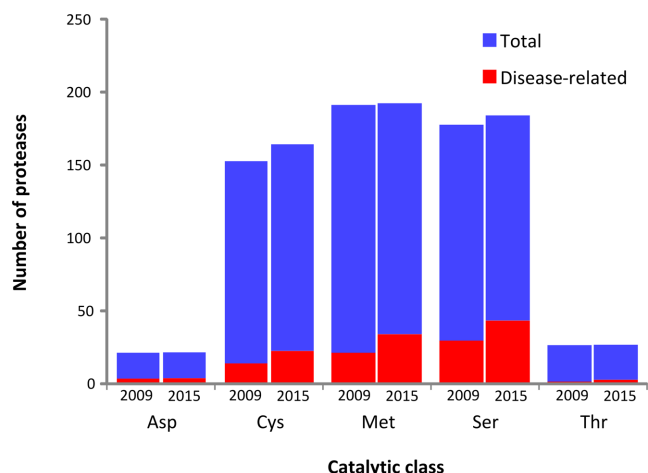


Figure 1. New annotations of proteases and degradomopathies. The number of proteases and degradomopathies annotated in each catalytic class is represented at the time of creation of the database (2009) and at the current version (2015).

evance of some proteases, which is not directly available in other databases. In our view, these databases taken together provide an accurate depiction of the current knowledge about degradomics.

In these last years, the field of degradomics has experienced a remarkable expansion, not so much in the number of known proteases as in the biological and pathological roles played by the degradome. Thus, almost all of the growth in the number of known proteases has developed through the inclusion of new protein families which were not previously known to display proteolytic activity. This suggests that the initial annotation of proteases in these model genomes was highly successful. However, the number of known degradomopathies has undergone a sharp increase (>50%) in the last six years. In this manuscript, we present the updates to the Degradome database, including new interactive representations which, in our opinion, provide a useful global depiction of the degradomes.

DATABASE ACCESS

Annotation of individual proteases

The overall organization of the database remains unchanged, with information about five catalytic classes (aspartyl-, cysteine-, metallo-, serine- and threonine-proteases) encompassing 82 protease families in four species. Compared to the previous version of this database, we have annotated seven additional families with 18 new human and chimpanzee proteases, 21 new mouse proteases and 17 new rat proteases (Figure 1). This growth reflects a body of biochemical research by multiple laboratories which has uncovered previously unknown proteolytic activities in known proteins (17–19).

The first mode of access is contained in five tables, one for each catalytic class. Each table displays the names of the protease families using the MEROPS classification system, the name of each protease, and the gene symbol for the protease in each species (Figure 2A). The table directly repre-

sents the orthology between proteases in different species, as well as the pseudogenes for which a functional ortholog exists in at least one of the selected mammalian species. Clicking each table cell gives the user access to additional information. Thus, the family name leads to a summary containing selected publications about that family of proteases, and, when available, a description of its structural features. In addition, protease-specific links open *popup* tables with information about the status and activity of each protease and, if available, hyperlinks to other databases, including MEROPS (16), NCBI Gene (20) and Ensembl (21). These tables also show if the protease is involved in a degradomopathy, along with a link to the OMIM database entry describing the disease.

Hereditary diseases of proteolysis

The information about mutated proteases in hereditary diseases is kept into a separate table (<http://degradome.uniovi.es/diseases.html>), to bypass the need to browse or search the individual annotations. This table of degradomopathies contains information about causal gene locus, mode of inheritance, pathologic protease alteration (gain/loss of proteolytic activity), and availability of described animal models containing the same protease anomaly. A link to related OMIM entries is also provided.

To our knowledge, this remains the only summary of the relationships between degradomics and pathology. In six years, the number of entries has grown from 77 to 119, reflecting the intense research into the pathological implications of the degradome. In fact, several proteases, such as ADAM10 and AFG3L2, have been related to more than one hereditary disease through different alterations. This table does not reflect the whole contribution of the degradome to human disease, as it does not encompass the numerous examples of non-hereditary diseases in which proteases are known to play an important role through alterations in their spatio-temporal patterns of expression. In this regard, the Degradome database has also demonstrated its usefulness in the analysis of proteases associated with cancer (22,23) and ageing (4,24).

Graphic interface

Ever since its definition, a recurrent global representation of the degradome has featured a characteristic circular drawing (Figure 2B). In this new version of the Degradome database, we have developed a tool that depicts the whole database as an svg figure. In addition to being customizable and constantly updated, this format allows a fair level of interaction with the user. Thus, each individual protease object contains a hyperlink to the corresponding annotation table. This figure can also be modified to represent the results of a search.

More importantly, this figure provides a global view of the degradomes of the included species. This not only provides a sense of the sizes of protease families, but also a fast and intuitive comparison between the degradomes of different species, including family expansions (i.e. mouse and rat expansions of the S01 family) and pseudogenization events. This interactive figure is used at this time in the home page

A

Family	Name	Human	Chimpanzee	Mouse	Rat
	napsin B	NAPSB	NAPSB		
	submandibular renin			Ren2	
A02	DDI-related protease	ASPRV1	DDI-RP	Asprv1	Asprv1
	DNA-damage inducible protein	DDI1	DDI1	Ddi1	Ddi1
	DNA-damage inducible protein 2	DDI2	DDI2	Ddi2	Ddi2
	nuclear recept. interacting prot. 2	NRIP2	NRIP2	Nrip2	Nrip2
	nuclear recept. interacting prot. 3	NRIP3	NRIP3	Nrip3	Nrip3
A22	presenilin 1	PSEN1	PSEN1	Psen1	Psen1
	presenilin 2	PSEN2	PSEN2	Psen2	Psen2
	presenilin homolog 1/SP	PSEN1IP1	PSEN1IP1	Psen1ip1	Psen1ip1
	presenilin homolog 2	PSEN1IP2	PSEN1IP2	Psen1ip2	Psen1ip2
	presenilin homolog 3/SP	PSEN1IP3	PSEN1IP3	Psen1ip3	Psen1ip3
	presenilin homolog 4/SPPL2B	PSEN1IP4	PSEN1IP4	Psen1ip4	Psen1ip4
	presenilin homolog 5	PSEN1IP5	PSEN1IP5	Psen1ip5	Psen1ip5
Ax1	GCDFP15	GCDFP15	GCDFP15	GCDFP15	GCDFP15
	seminal vesicle antigen	SVL1	SVL1	Sval1	Sval1
	seminal vesicle antigen-like 1	SVL2	SVL2	Sval2	Sval2

B

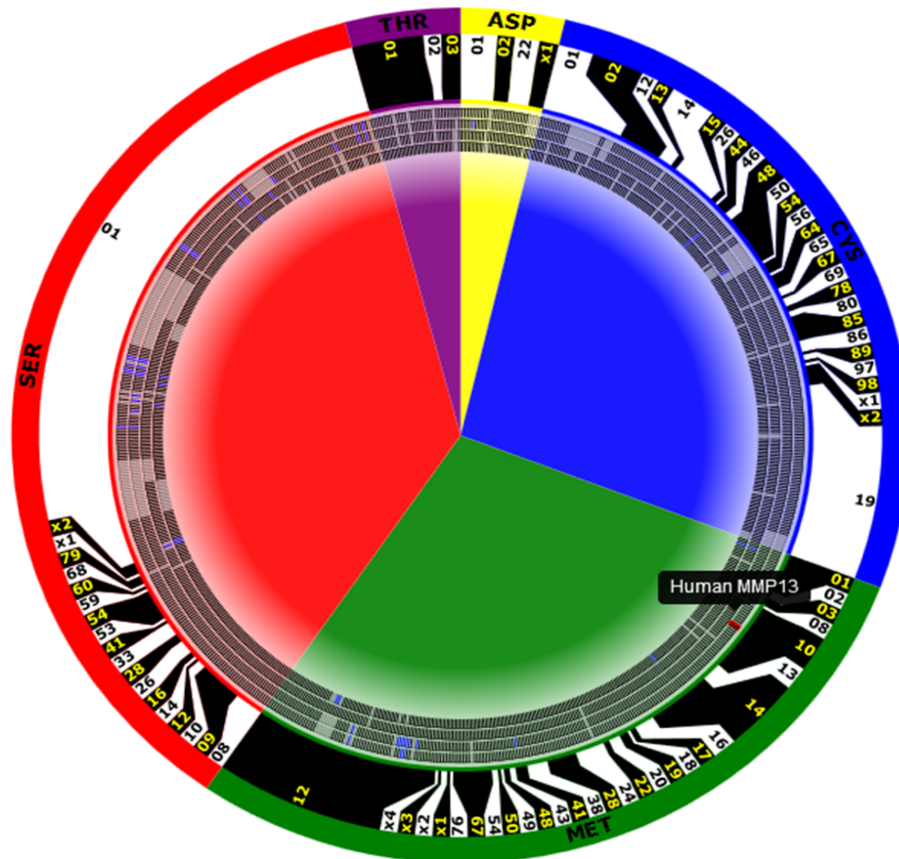


Figure 2. New features of the Degradome database. (A) Individual annotations of aspartyl proteases. The first column contains the name of the family, with a hyperlink to a web page where the user can find related selected publications and structures. The second column contains the name of each protease, with a hyperlink which opens a *popup* table with further general information—in this example, *presenilin 1*. Pseudogenes are shown over a *pink background*. Proteases absent in a species are shown as *empty grey cells*. (B) Interactive representation of the degradomes of (from outer to inner ticks) human, chimpanzee, mouse and rat. Protease families are limited with *black* and *white boxes*. Catalytic classes are shown as background *colored arcs*. Proteases which have been pseudogenized are depicted as *blue ticks*, and proteases absent in an organism are shown as *grey ticks*. Human collagenase-3 is highlighted to show the interactivity of the ticks.

(<http://degradome.uniovi.es/dindex.html>), as an alternative entry method for the database, and as a snapshot of the differences between the degradomes of human and mouse (<http://degradome.uniovi.es/hmd.html>).

Additional contents

In addition to the Degradome database, the web site also keeps offering several summaries of the characteristics of mammalian degradomes. Thus, a static table listing human, mouse and rat protease inhibitors can be found at <http://degradome.uniovi.es/inhibitors.html>. A count of proteases in these species, itemized by catalytic class, is shown at <http://degradome.uniovi.es/numbers.html>. These numbers are kept updated as novel catalytic classes are discovered and added to the Degradome database. Additionally, we also keep a figure showing the different ancillary domains present in proteases (<http://degradome.uniovi.es/domains.html>). Interactive structures for different protease families are also kept in pdf format for teaching purposes (<http://degradome.uniovi.es/structures.html>).

IMPLEMENTATION

The database with the annotations of individual proteases has been migrated to a single JSON file, which is freely available upon request. The information is queried from the web interface using the AJAX technology through JQuery. Therefore, if the browser lacks Javascript or Javascript is blocked, the user is offered a link to a static table displaying all of the information at once. The style of the web pages is implemented with the Bootstrap library to increase accessibility from multiple devices. The new graphical interface is written in svg, which is directly generated from the degradome JSON file using a custom Perl script.

Finally, *selected structures* are displayed in pdf format. Thus, the user needs Adobe Reader v7.0 or higher. Several reasons may hamper the viewing for these files from the common browsers. If this happens, the user can download the pdf file and access its contents locally.

CONCLUSION AND FUTURE DIRECTION

The Degradome database has grown in the last years reflecting the advances in our understanding of proteases in biological and pathological processes. As a part of our involvement in degradomics, we will continue updating this database as more results become available. Based on our experience, we expect this increase to continue, driven mainly by two sources: biochemical studies that uncover novel proteolytic activities and high-throughput association studies which relate additional protease mutations with hereditary diseases.

FUNDING

European Union [CancerDegradome-FP6 and MicroEnviMet-FP7]; Ministerio de Economía y Competitividad-Spain, Principado de Asturias and Fundación Botín (supported by Banco Santander through its Santander Universities Global Division). Funding for open access charge: Ministerio de Economía y Competitividad-Spain.

Conflict of interest statement. None declared.

REFERENCES

- Lopez-Otin, C. and Overall, C.M. (2002) Protease degradomics: a new challenge for proteomics. *Nat. Rev. Mol. Cell. Biol.*, **3**, 509–519.
- Lopez-Otin, C. and Bond, J.S. (2008) Proteases: multifunctional enzymes in life and disease. *J. Biol. Chem.*, **283**, 30433–30437.
- Reinhard, S.M., Razak, K. and Ethell, I.M. (2015) A delicate balance: role of MMP-9 in brain development and pathophysiology of neurodevelopmental disorders. *Front. Cell Neurosci.*, **9**, 280.
- Quiros, P.M., Langer, T. and Lopez-Otin, C. (2015) New roles for mitochondrial proteases in health, ageing and disease. *Nat. Rev. Mol. Cell. Biol.*, **16**, 345–359.
- Voskoboinik, I., Whisstock, J.C. and Trapani, J.A. (2015) Perforin and granzymes: function, dysfunction and human pathology. *Nat. Rev. Immunol.*, **15**, 388–400.
- Turk, B., Turk, D. and Turk, V. (2012) Protease signalling: the cutting edge. *EMBO J.*, **31**, 1630–1643.
- Nalivaeva, N.N., Belyaev, N.D., Kerridge, C. and Turner, A.J. (2014) Amyloid-clearing proteins and their epigenetic regulation as a therapeutic target in Alzheimer's disease. *Front. Aging Neurosci.*, **6**, 235.
- Gordon, L.B., Rothman, F.G., Lopez-Otin, C. and Misteli, T. (2014) Progeria: a paradigm for translational medicine. *Cell*, **156**, 400–407.
- Gutierrez-Fernandez, A., Soria-Valles, C., Osorio, F.G., Gutierrez-Abril, J., Garabaya, C., Aguirre, A., Fueyo, A., Fernandez-Garcia, M.S., Puente, X.S. and Lopez-Otin, C. (2015) Loss of MT1-MMP causes cell senescence and nuclear defects which can be reversed by retinoic acid. *EMBO J.*, **34**, 1875–1888.
- Lopez-Otin, C. and Hunter, T. (2010) The regulatory crosstalk between kinases and proteases in cancer. *Nat. Rev. Cancer*, **10**, 278–292.
- Puente, X.S., Sanchez, L.M., Overall, C.M. and Lopez-Otin, C. (2003) Human and mouse proteases: a comparative genomic approach. *Nat. Rev. Genet.*, **4**, 544–558.
- Puente, X.S., Sanchez, L.M., Gutierrez-Fernandez, A., Velasco, G. and Lopez-Otin, C. (2005) A genomic view of the complexity of mammalian proteolytic systems. *Biochem. Soc. Trans.*, **33**, 331–334.
- Ordóñez, G.R., Puente, X.S., Quesada, V. and Lopez-Otin, C. (2009) Proteolytic systems: constructing degradomes. *Methods Mol. Biol.*, **539**, 33–47.
- Quesada, V., Ordóñez, G.R., Sanchez, L.M., Puente, X.S. and Lopez-Otin, C. (2009) The Degradome database: mammalian proteases and diseases of proteolysis. *Nucleic Acids Res.*, **37**, D239–243.
- Igarashi, Y., Heures, E., Doctor, K.S., Talwar, P., Gramatikova, S., Gramatikoff, K., Zhang, Y., Blinov, M., Ibragimova, S.S., Boyd, S. *et al.* (2009) PMAP: databases for analyzing proteolytic events and pathways. *Nucleic Acids Res.*, **37**, D611–D618.
- Rawlings, N.D., Waller, M., Barrett, A.J. and Bateman, A. (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.*, **42**, D503–D509.
- Bujalka, H., Koenning, M., Jackson, S., Perreau, V.M., Pope, B., Hay, C.M., Mitew, S., Hill, A.F., Lu, Q.R., Wegner, M. *et al.* (2013) MYRF is a membrane-associated transcription factor that autoproteolytically cleaves to directly activate myelin genes. *PLoS Biol.*, **11**, e1001625.
- Liang, J., Saad, Y., Lei, T., Wang, J., Qi, D., Yang, Q., Kolattukudy, P.E. and Fu, M. (2010) MCP-induced protein 1 deubiquitinates TRAF proteins and negatively regulates JNK and NF-kappaB signaling. *J. Exp. Med.*, **207**, 2959–2973.
- Stingle, J., Habermann, B. and Jentsch, S. (2015) DNA-protein crosslink repair: proteases as DNA repair enzymes. *Trends Biochem. Sci.*, **40**, 67–71.
- Brown, G.R., Hem, V., Katz, K.S., Ovetsky, M., Wallin, C., Ermolaeva, O., Tolstoy, I., Tatusova, T., Pruitt, K.D., Maglott, D.R. *et al.* (2015) Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res.*, **43**, D36–42.
- Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S. *et al.* (2015) Ensembl 2015. *Nucleic Acids Res.*, **43**, D662–D669.

22. Fraile, J.M., Quesada, V., Rodriguez, D., Freije, J.M. and Lopez-Otin, C. (2012) Deubiquitinases in cancer: new functions and therapeutic options. *Oncogene*, **31**, 2373–2388.
23. Sevenich, L. and Joyce, J.A. (2014) Pericellular proteolysis in cancer. *Genes Dev.*, **28**, 2331–2347.
24. Fernandez, A.F. and Lopez-Otin, C. (2015) The functional and pathologic relevance of autophagy proteases. *J. Clin. Invest.*, **125**, 33–41.